STATISTICAL ANALYSIS PLAN (SAP)

Home monitoring of patients with rheumatoid arthritis - an eHealth development study: A SAP for a Prospective cohort study

Version: July 26, 2017

COLLABORATORS:

Marie Skougaard, Sabrina M. Nielsen, Robin Christensen, Henning Bliddal, and others involved in the ELECTOR project

The trial was registered online at clinicaltrials.gov (NCT02317939)
Link here: https://clinicaltrials.gov/ct2/show/NCT02317939

Statistical Analyst:
Sabrina M. Nielsen, MSc
Biostatistics and Clinical Epidemiology fellow
Musculoskeletal Statistics Unit, The Parker Institute,
Bispebjerg and Frederiksberg Hospital,
DK-2000 Copenhagen F, Denmark

Statistical Advisor:
Dr. Robin Christensen, MSc, PhD; Biostatistician
Professor of Clinical Epidemiology, adj.
Head of Musculoskeletal Statistics Unit,
The Parker Institute, Bispebjerg and Frederiksberg Hospital,
DK-2000 Copenhagen F, Denmark

PURPOSE OF THE STATISTICAL ANALYSIS PLAN (SAP)

The SAP is intended to bring the team together on the same page.
The SAP adds another layer of specificity to the project; it was prepared after collecting the data, but before conducting any analyses.

The SAP was written and completed after finalizing the protocol. In this document, a more technical and detailed elaboration of the principal features stated in the protocol is included. The plan includes detailed procedures for executing the statistical analyses of the primary and secondary variables and other data.

INTRODUCTION

Telemonitoring and eHealth solutions for assessing patients with chronic illnesses as diabetes, asthma and hypertension have previously shown great advantages in better control of illnesses and improvement of symptoms (1). A similar eHealth solution for patients with rheumatoid arthritis (RA) is expected to be advantageous both for patients and the health care system.

According to the European League Against Rheumatism (EULAR) recommendations, tight control of disease activity is necessary to ensure optimal treatment of RA (2, 3). Currently, this tight control is managed in the clinic by physicians, nurses, and biometricians, which is expensive and time consuming for both patients and health care professionals (HCPs)(4, 5). The possibility for self-management of stable patients as a part of an eHealth solution would leave more time in the clinic for patients in specific need of care and provide greater involvement of patients themselves, factors which have proven health benefits (6).

Self-management and involvement of patients will be a central part of an eHealth solution. This necessitates reliability of patient-performed assessments including both patient reported outcomes via questionnaires and evaluation of swollen and tender joints. The reliability of these assessments will be crucial to a future eHealth solution and fundamental for reaching optimal monitoring and treatment.

Test-retest reliability of the disease activity score (DAS28) has previously been assessed in different studies (4, 7-16) However, the problem of poor agreement as well as intra- and inter-observer variations in assessment of swollen joint especially, is indeed present (10, 11). Issues regarding the effect of training have been disclosed with inconclusive results. A study of 30 RA patients with stable disease found no definite value of 10 minutes of training (10). By principle, instruction in distinction between inflammatory and non-inflammatory joint swelling (soft vs. bony swellings) would be expected to be useful, while no evidence was described of more valid assessment made by patients receiving formal joint count training compared to patients who received no training (14). By consequence, the effect of training in relation to the reliability of patients' joint counts remains to be clarified.

To date, no sufficiently powered study has yet compared assessments performed by patients in a study simulating a real-life setup with follow-up visits. This study aims to assess the intra- and interrater reliability of patients and compared to HCPs at baseline and at follow-up.

We hypothesize that acceptable reliability of patients' self-performed joint assessments may be obtained with a simple instruction program and is sufficient for the use of patient derived data via an eHealth solution that will be a part of the future health care setup for RA patients.

DATA SOURCE

Data originate from a two-step multi-centre randomised controlled trial with four visits (T1, T2, T3, T4). The trial was a part of the ELECTOR project (www.elector.eu) where an online platform had been constructed for patients reporting and filling in questionnaires from home. All data was collected in a to the platform connected database.

Patient inclusion criteria were: diagnosed with RA ≥ 12 months, DAS28CRP ≤ 5.1, and age between 18 and 85 years. Patient exclusion criteria were: dementia or other linguistic/cognitive/physical deficiency that prevents participation, vision impairment that prevents the use of the devices and computer.

At baseline (T1), all patients participating in the study were provided with a single instruction and training session of 15 minutes. Next, patients were randomized to one of the following four groups, 1A, 1B, 2A, 2B. '1' means that the patients were assessed only by themselves, whereas '2' means that the patients additionally were assessed by HCPs and US. 'A' means that the patients a received training refreshment via online instruction prior to the assessment performed at T3. 'B' means no training refreshment.

The study was conducted in Prague and Copenhagen. In Prague, patients were only randomized to 1A and 1B.

T1 will refer to the assessment after training has been received, unless otherwise stated.

ANALYSIS OBJECTIVES

The main analysis objective is to investigate the intrarater reliability and agreement for DAS28CRP done by patients at two visits placed close enough that a change in disease activity is not anticipated, i.e. T1 and T2.

Secondary analysis objectives include:

(a) Investigate the interrater reliability and agreement for DAS28CRP done by HCPs and patients at T1.

(b) Investigate the intra- and interrater reliability and agreement for DAS28CRP when stratifying according to amount of training (one or two times) at T3 vs T4 and at T3, respectively.

(c) Investigate the intra- and interrater reliability for assessing tender joints and swollen joints, i.e. done by patients at two visits placed close enough that a change in disease activity is not anticipated, i.e. at T1 and T2, and done by two HCPs, patients, and by ultrasound (US) examination at T1.

(d) Investigate the intrarater reliability and agreement for DAS28CRP done by patients at T1 before vs after initial training.

ANALYSIS SETS/ POPULATIONS/SUBGROUPS

The intention-to-treat (ITT), complete cases (CC), as observed (AsO) and per protocol (PP) populations will be defined according to the following criteria:

ITT:

- Patients, who were eligible for inclusion
- Patients, who were randomized

AsO for the outcome of interest:

- Patients, who has data on the outcome

CC for outcome of interest:

- Patients, who have complete data for the analysis on the outcome of interest at the visit(s) of interest for all possible raters (for intra-rater reliability raters include only patients)

PP:

- Patients, who have complete data on VAS and joint count at all visits and for all possible raters
- Visit 2 and 3 needs to be within 1-3 and 48-62 days, respectively, after visit 1, and visit 4 needs to be within 1-3 days after visit 3.

Our primary analysis population will be the PP population, whereas the other populations will serve as sensitivity analyses.

VARIABLES

DAS28CRP is calculated from the formula, $DAS28CRP = 0.56 \cdot \sqrt{(TJC28)} + 0.28 \cdot \sqrt{(SJC28)} + 0.36 \cdot \ln(CRP+1) + 0.014 \cdot GH + 0.96$, by M. FLendrie and J. Fransen (https://www.das-score.nl/das28/DAScalculators/dasculators.xls)where TJC28 is the number of tender joint, SJC28 is the number of swollen joints, CRP is the level of CRP, and GH is 'general health' which equals VAS global.

The CRP was only measured at T1 and T3, whereas the CRP for calculating DAS28CRP at T2 and T4 was the CRP measured at T1 and T3, respectively.

Swollen joint and tender joints were assessed for each of 28 joints by the patients or HCPs, and for each of 12 joints by a US examinator. Patients and HCPs assessed by palpating of the joints. The US examinator assessed presence of

Doppler activity and synovial hypertrophia. Swollen joints are considered related to synovial hypertrophia and Doppler activity is considered related to tender joints.

HANDLING OF MISSING VALUES AND OTHER DATA CONVENTIONS

For the ITT population, missing data on DAS28CRP will be imputed using grand mean. For swollen and tender joints, no imputation of missing data will be conducted.

STATISTICAL PROCEDURES

The characteristics of the participants will be described presenting binary outcomes as numbers with corresponding percentages for categorical data, continuous outcomes as means with corresponding standard deviations (SD), and ordinal outcomes (or continuous data that are not normally distributed) as medians with corresponding interquartile range (IQR).

For the quantification of intra- and inter-rater reliability for the DAS28CRP score, the intraclass correlation coefficient (ICC) will be applied (17). ICC ranges between 0 (no reliability) and 1 (perfect reliability). Values <0.40 will be interpreted as poor reliability, 0.40 to 0.59 as fair reliability, 0.60 to 0.74 as good reliability and 0.75-1.00 as excellent reliability (18). The ICC will be of the type ICC(2,1), i.e. based on a two-way random effects model estimating agreement with single measures (17, 19)).

Intra- and interrater agreement will be assessed using Bland-Altman plots. The Bland-Altman plot provides insight into the distribution of differences in relation to mean values (20). The limits of agreement will be calculated as $\pm 1.96 \cdot SD$, where the SD is the SD of the differences. If the differences within the limits of agreement is not considered clinically important, the two sets of ratings (e.g. time 1 vs time 2, or rater 1 vs rater 1) can be considered interchangeably (20). Non-important difference in DAS28CRP will be defined as 0.6.

The coefficient of variation (CV) will be calculated as $SD/mean \cdot 100$, and measures the relative variability. The CV will be calculated for each patient at each visit and data will be shown as a figure by plotting CV% against visit for each patient.

The minimal detectable difference (MD) will be calculated from $SEP \cdot 1.96 \cdot \sqrt{2}$, where SEP will be calculated from $SD \cdot \sqrt{(1-ICC)}$, and here the SD is the SD of all scores from all subjects (Joseph P. Weir 2005). The MD represents the minimal difference that must be shown to ensure that the observed difference is real and not just measurement error.

For the quantification of intra- and interrater reliability for the classification of tender and swollen joints, we will use Cohen’s Kappa coefficients (21). Kappa estimates will be interpreted according to Landis and Koch (1977)(22) where <0 is poor agreement, 0.0 – 0.20 is slight agreement, 0.21 – 0.40 is fair agreement, 0.41 – 0.60 is moderate agreement, 0.61 – 0.80 is substantial agreement, and 0.81 – 1.00 is almost perfect agreement.

Furthermore, the observed agreement will be reported, i.e. number of patients for which both raters are agreeing.

Assumptions for the analyses will be checked, such as the differences follow a normal distribution when calculating limits of agreement. All calculations will be carried out using the statistical software R (version 3.3.3 or newer) (23) with the package “psych” (24).

Program code for R (version 3.3.3 or newer)

```
#Read and prepare data
Data=read.table("clipboard",header=TRUE)
DataT1<-subset(Data,Data$Visit=="1")
DataT2<-subset(Data,Data$Visit=="2")
DataICC<-cbind(DataT1$DAS28CRP_PT,DataT2$DAS28CRP_PT)
DataKappa<-cbind(DataT1$TJ1_PT,DataT2$TJ1_PT)

#ICC(2,1)
psych::ICC(DataICC)

#Cohen's Kappa
psych::cohen.kappa(DataKappa)
```

MEASURES TO ADJUST FOR MULTIPLICITY

In all analyses, p-values < 0.5 will be considered statistically significant. No correction for multiple testing will be done; instead the results will be interpreted with caution.

SENSITIVITY ANALYSES

In order to evaluate the robustness of the results, sensitivity analyses according to different analysis populations will be carried out as described in the manuscript outline. Furthermore, intrarater analyses will be conducted separately for Prague and Copenhagen.

PROPOSED DATASET STRUCTURE

PtID	Age	Gender	DiseaseDuration	csDMARDuse	bDMARDuse	PrednisoloneUse	VAS global	VAS pain	VAS fatigue	Centre	ExtraMonitoring(1/0)	ExtraTraining(A/B)	Visit	Day from baseline	DAS28CRP_PT	DAS28CRP_HCP1	DAS28CRP_HCP2	TJ1_PT	TJ2_PT	TJ3_PT	TJ4_PT	TJ5_PT	TJ6_PT	TJ7_PT	TJ8_PT	TJ9_PT	TJ10_PT	TJ11_PT	TJ12_PT	TJ13_PT	TJ14_PT	TJ15_PT	TJ16_PT	TJ17_PT	TJ18_PT	TJ19_PT	TJ20_PT	TJ21_PT	TJ22_PT	TJ23_PT	TJ24_PT	TJ25_PT	...
1													0																														...
1													1																														...
1													2																														...
1													3																														...
1													4																														...
2													1																														...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

...	TJ26_PT	TJ27_PT	TJ28_PT	TJ1_HCP1	TJ2_HCP1	TJ3_HCP1	TJ4_HCP1	TJ5_HCP1	TJ6_HCP1	TJ7_HCP1	TJ8_HCP1	TJ9_HCP1	TJ10_HCP1	TJ11_HCP1	TJ12_HCP1	TJ13_HCP1	TJ14_HCP1	TJ15_HCP1	TJ16_HCP1	TJ17_HCP1	TJ18_HCP1	TJ19_HCP1	TJ20_HCP1	TJ21_HCP1	TJ22_HCP1	TJ23_HCP1	TJ24_HCP1	TJ25_HCP1	TJ26_HCP1	TJ27_HCP1	TJ28_HCP1	TJ1_HCP2	TJ2_ HCP2	TJ3_ HCP2	TJ4_ HCP2	TJ5_ HCP2	TJ6_HCP2	TJ7_HCP2	TJ8_HCP2	TJ9_HCP2	TJ10_HCP2	TJ11_HCP2	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

...	TJ12_HCP2	TJ13_HCP2	TJ14_HCP2	TJ15_HCP2	TJ16_HCP2	TJ17_HCP2	TJ18_HCP2	TJ19_HCP2	TJ20_HCP2	TJ22_HCP2	TJ23_HCP2	TJ24_HCP2	TJ25_HCP2	TJ26_HCP2	TJ27_HCP2	TJ28_HCP2	TJ1_US	TJ2_ US	TJ3_ US	TJ4_ US	TJ5_ US	TJ6_ US	TJ7_US	TJ8_ US	TJ9_ US	TJ10_ US	TJ11_ US	TJ12_ US	SJ1_PT	SJ2_ PT	SJ3_ PT	SJ4_ PT	SJ5_ PT	SJ6_ PT	SJ7_ PT	SJ8_ PT	SJ9_ PT	SJ10_ PT	SJ11_ PT	SJ12_ PT	SJ13_ PT	SJ14_ PT	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

...	SJ15_PT	SJ16_PT	SJ17_PT	SJ18_PT	SJ19_PT	SJ20_PT	SJ21_PT	SJ22_PT	SJ23_PT	SJ24_PT	SJ25_PT	SJ26_PT	SJ27_PT	SJ28_PT	SJ1_HCP1	SJ2_ HCP1	SJ3_ HCP1	SJ4_ HCP1	SJ5_HCP1	SJ6_HCP1	SJ7_HCP1	SJ8_HCP1	SJ9_HCP1	SJ10_HCP1	SJ11_HCP1	SJ12_HCP1	SJ13_HCP1	SJ14_HCP1	SJ15_HCP1	SJ16_HCP1	SJ17_HCP1	SJ18_HCP1	SJ19_HCP1	SJ20_HCP1	SJ21_HCP1	SJ22_HCP1	SJ23_HCP1	SJ24_HCP1	SJ25_HCP1	SJ26_HCP1	SJ27_HCP1	SJ28_HCP1	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

...	SJ1_HCP2	SJ2_ HCP2	SJ3_ HCP2	SJ4_ HCP2	SJ5_HCP2	SJ6_HCP2	SJ7_HCP2	SJ8_HCP2	SJ9_HCP2	SJ10_HCP2	SJ11_HCP2	SJ12_HCP2	SJ13_HCP2	SJ14_HCP2	SJ15_HCP2	SJ16_HCP2	SJ17_HCP2	SJ18_HCP2	SJ19_HCP2	SJ20_HCP2	SJ21_HCP2	SJ22_HCP2	SJ23_HCP2	SJ24_HCP2	SJ25_HCP2	SJ26_HCP2	SJ27_HCP2	SJ28_HCP2	SJ1_US	SJ2_ US	SJ3_ US	SJ4_ US	SJ5_US	SJ6_US	SJ7_US	SJ8_ US	SJ9_US	SJ10_US	SJ11_US	SJ12_US	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

PROPOSED MANUSCRIPT OUTLINE (FIGURES, TABLES, APPENDICES)

Figure 1: Flow diagram

Table 1: Baseline characteristics

Table 2: Intra- and interrater reliability DAS28CRP (per protocol)

Figure 2a-d: Intrarater agreement in patient ratings between visit 1 and 2 [primary], pre- and post-training at visit 1, and visit 3 and 4 for subgroup A and B, using Bland-Altman plots (per protocol)

Figure 3a-c: Interrater agreement between Pt-HCP1, Pt-HCP2, HCP1-HCP2, respectively, using Bland-Altman plots, at visit 1 (per protocol)

Figure 4: Interrater agreement between Pt, HCP1, and HCP2, using coefficient of variation for all visits (per protocol)

Table 3: Intrarater reliability tender and swollen joints (per protocol), visit 1

Table 4: Interrater reliability tender joints (per protocol), visit 1

Table 5: Interrater reliability swollen joints (per protocol), visit 1

APPENDIX

Table 2: Intra- and interrater reliability DAS28crp (ITT, grand mean imputation) *

Table 2: Intra- and interrater reliability DAS28crp (as observed) *

Table 2: Intra- and interrater reliability DAS28crp (complete cases) *

Table 2: Intrarater reliability DAS28crp stratified according to centre *

Figure 2a-d: Intrarater agreement in patient ratings between visit 1 and 2 [primary], pre- and post-training at visit 1, and visit 3 and 4 for subgroup A and B, using Bland-Altman plots (complete cases) *

Figure 3a-c: Interrater agreement between Pt-HCP1, Pt-HCP2, HCP1-HCP2, respectively, using Bland-Altman plots, at visit 1 (complete cases) *

Figure 3a-b: Interrater agreement between Pt-HCP1, Pt-HCP2, respectively, using Bland-Altman plots, at visit 1 pre-training (per protocol)

Figure 3a-b: Interrater agreement between Pt-HCP1, Pt-HCP2, respectively, using Bland-Altman plots, at visit 3, subgroup A (per protocol)

Figure 3a-b: Interrater agreement between Pt-HCP1, Pt-HCP2, respectively, using Bland-Altman plots, at visit 3, subgroup B (per protocol)

*Sensitivity analyses

Figure 1: Flow diagram

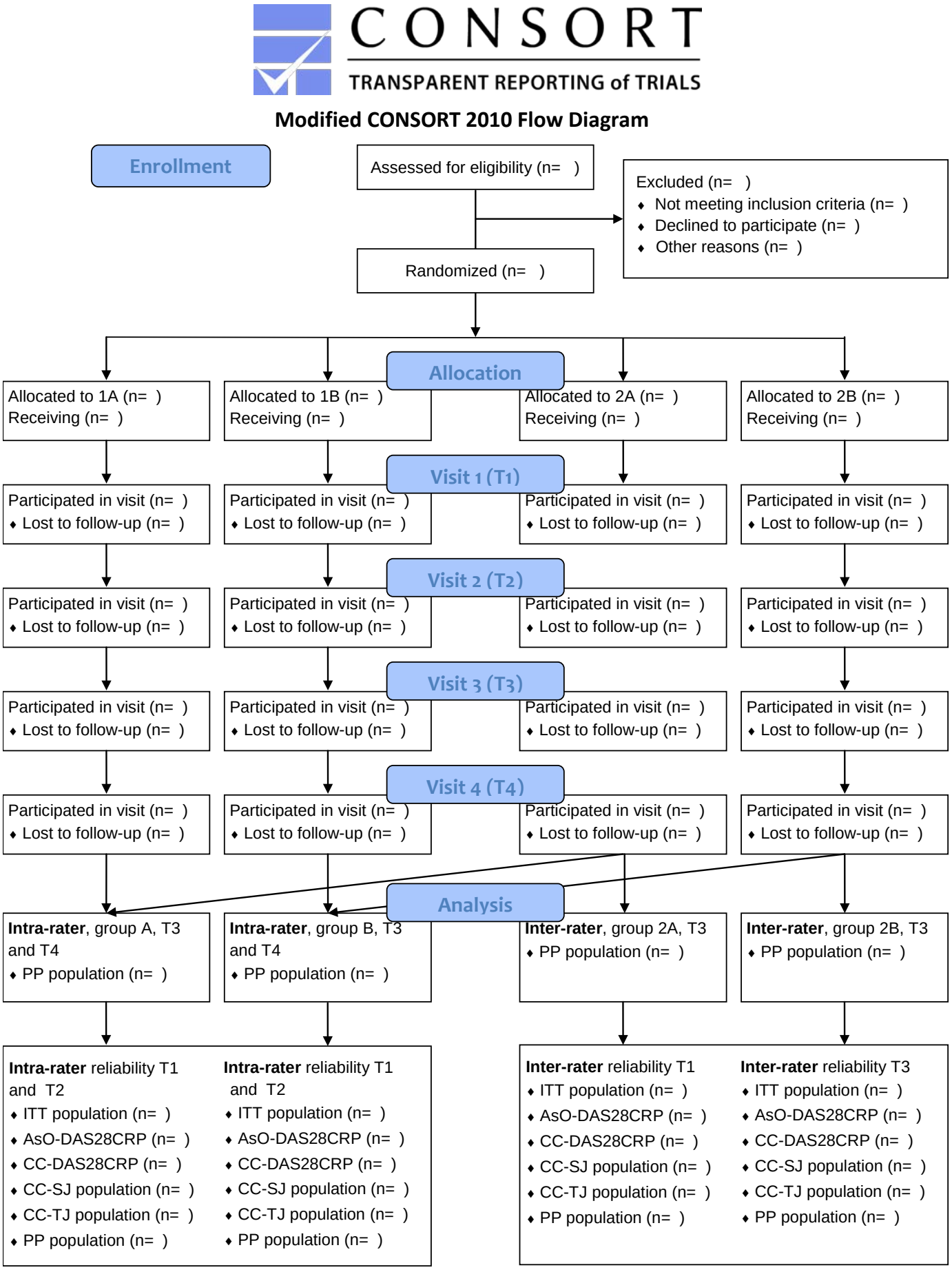

Table 1: Baseline characteristics

Demographics	Patients PP (n=XX)	Patients non-PP (n=XX)
Age, years	XX.X (XX.X)	XX.X (XX.X)
Female, n (%)	XX (XX%)	XX (XX%)
Disease duration, years	XX.X (XX.X)	XX.X (XX.X)
csDMARD use, n (%)	XX (XX%)	XX (XX%)
bDMARD us, n (%)	XX (XX%)	XX (XX%)
Prednisolone use, n (%)	XX (XX%)	XX (XX%)
DAS28CRP	XX.X (XX.X)	XX.X (XX.X)
VAS global	XX.X (XX.X)	XX.X (XX.X)
VAS pain	XX.X (XX.X)	XX.X (XX.X)
VAS fatigue	XX.X (XX.X)	XX.X (XX.X)

Data is presented as mean with corresponding standard deviations unless other is stated.

Table 2: Intra- and interrater reliability DAS28CRP (per protocol)

	Intrarater reliability			Interrater reliability			
Variable	Number of patients, n	ICC (95%CI)	MD	Number of patients, n	Pt vs. HCP1 ICC (95%CI)	Pt vs. HCP2 ICC (95%CI)	HCP1 vs HCP2 ICC (95%CI)
Visit 1†, all [primary]	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Visit 3†, all	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Visit 3†, subgroup A	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Visit 3†, subgroup B	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Visit 1 pre-training‡, all	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	-

ICC, intraclass correlation coefficient; MDC, minimal detectable change. etc.

† For intrarater reliability, visit 1 and 2 are compared to visit 3 and 4, respectively.

‡ For intrarater reliability, visit 1 pre-training is compared to visit 1 post-training.

[figure 2a-d]

Figure 2a-d: Intrarater agreement in patient ratings between visit 1 and 2 [primary], pre- and post-training at visit 1, and visit 3 and 4 for subgroup A and B, using Bland-Altman plots (per protocol)

[figure 3a-c]

Figure 3a-c: Interrater agreement between Pt-HCP1, Pt-HCP2, HCP1-HCP2, respectively, using Bland-Altman plots, at visit 1 (per protocol)

[figure 4]

Figure 4: Interrater agreement between Pt, HCP1, and HCP2, using coefficient of variation for all visits (per protocol)

Table 3: Intrarater reliability tender and swollen joints (per protocol), visit 1

		Swollen joints			Tender joints		
Variable	Number of patients, n	Absolute agreement (%)	Kappa (95%CI)	MD	Absolute agreement (%)	Kappa (95%CI)	SDC
Joint 1	XX	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 2	XX	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 3	XX	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 4	XX	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 5	XX	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 6	XX	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 7	XX	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 8	XX	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 9	XX	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 10	XX	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 11	XX	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 12	XX	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 13	XX	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 14	XX	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 15	XX	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 16	XX	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 17	XX	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 18	XX	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 19	XX	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 20	XX	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 21	XX	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 22	XX	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 23	XX	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 24	XX	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 25	XX	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 26	XX	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 27	XX	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 28	XX	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	XX (XX%)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)

Table 4: Interrater reliability tender joints (per protocol), visit 1

Variable	Number of patients, n	Pt vs. HCP1 Kappa (95%CI)	Pt vs. HCP2 Kappa (95%CI)	HCP1 vs HCP2 Kappa (95%CI)	Pt vs. US Kappa (95%CI)	HCP1 vs. US Kappa (95%CI)	HCP2 vs. US Kappa (95%CI)
Joint 1	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 2	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 3	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 4	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 5	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 6	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 7	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 8	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 9	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 10	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 11	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 12	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 13	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	-	-	-
Joint 14	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	-	-	-
Joint 15	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	-	-	-
Joint 16	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	-	-	-
Joint 17	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	-	-	-
Joint 18	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	-	-	-
Joint 19	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	-	-	-
Joint 20	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	-	-	-
Joint 21	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	-	-	-
Joint 22	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	-	-	-
Joint 23	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	-	-	-
Joint 24	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	-	-	-
Joint 25	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	-	-	-
Joint 26	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	-	-	-
Joint 27	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	-	-	-
Joint 28	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	-	-	-

Table 5: Interrater reliability swollen joints (per protocol), visit 1

Variable	Number of patients, n	Pt vs. HCP1 Kappa (95%CI)	Pt vs. HCP2 Kappa (95%CI)	HCP1 vs HCP2 Kappa (95%CI)	Pt vs. US Kappa (95%CI)	HCP1 vs. US Kappa (95%CI)	HCP2 vs. US Kappa (95%CI)
Joint 1	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 2	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 3	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 4	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 5	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 6	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 7	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 8	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 9	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 10	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 11	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 12	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)
Joint 13	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	-	-	-
Joint 14	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	-	-	-
Joint 15	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	-	-	-
Joint 16	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	-	-	-
Joint 17	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	-	-	-
Joint 18	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	-	-	-
Joint 19	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	-	-	-
Joint 20	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	-	-	-
Joint 21	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	-	-	-
Joint 22	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	-	-	-
Joint 23	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	-	-	-
Joint 24	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	-	-	-
Joint 25	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	-	-	-
Joint 26	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	-	-	-
Joint 27	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	-	-	-
Joint 28	XX	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	0.XX (0.XX to 0.XX)	-	-	-

References

1. Pare G, Moqadem K, Pineau G, St-Hilaire C. Clinical effects of home telemonitoring in the context of diabetes, asthma, heart failure and hypertension: a systematic review. Journal of medical Internet research. 2010;12(2):e21.
2. Knevel R, Schoels M, Huizinga TW, Aletaha D, Burmester GR, Combe B, et al. Current evidence for a strategic approach to the management of rheumatoid arthritis with disease-modifying antirheumatic drugs: a systematic literature review informing the EULAR recommendations for the management of rheumatoid arthritis. Annals of the rheumatic diseases. 2010;69(6):987-94.
3. Smolen JS, Landewe R, Breedveld FC, Buch M, Burmester G, Dougados M, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2013 update. Annals of the rheumatic diseases. 2014;73(3):492-509.
4. Cheung PP, Ruyssen-Witrand A, Gossec L, Paternotte S, Le Bourlout C, Mazieres M, et al. Reliability of patient self-evaluation of swollen and tender joints in rheumatoid arthritis: A comparison study with ultrasonography, physician, and nurse assessments. Arthritis care & research. 2010;62(8):1112-9.
5. Kirwan JR, Snow SM. Which patients see a rheumatologist? SWRACR (South Western Regional Advisory Committee on Rheumatology). British journal of rheumatology. 1991;30(4):285-7.
6. Osborne RH, Wilson T, Lorig KR, McColl GJ. Does self-management lead to sustainable health benefits in people with arthritis? A 2-year transition study of 452 Australians. The Journal of rheumatology. 2007;34(5):1112-7.
7. Barton JL, Imboden J, Graf J, Glidden D, Yelin EH, Schillinger D. Patient-physician discordance in assessments of global disease severity in rheumatoid arthritis. Arthritis care & research. 2010;62(6):857-64.
8. Figueroa F, Braun-Moscovici Y, Khanna D, Voon E, Gallardo L, Luinstra D, et al. Patient self-administered joint tenderness counts in rheumatoid arthritis are reliable and responsive to changes in disease activity. The Journal of rheumatology. 2007;34(1):54-6.
9. Hanly JG, Mosher D, Sutton E, Weerasinghe S, Theriault D. Self-assessment of disease activity by patients with rheumatoid arthritis. The Journal of rheumatology. 1996;23(9):1531-8.
10. Heegaard C, Dreyer L, Egsmose C, Madsen OR. Test-retest reliability of the disease activity score 28 CRP (DAS28-CRP), the simplified disease activity index (SDAI) and the clinical disease activity index (CDAI) in rheumatoid arthritis when based on patient self-assessment of tender and swollen joints. Clinical rheumatology. 2013;32(10):1493-500.
11. Houssien DA, Stucki G, Scott DL. A patient-derived disease activity score can substitute for a physician-derived disease activity score in clinical research. Rheumatology (Oxford, England). 1999;38(1):48-52.
12. Kavanaugh A, Lee SJ, Weng HH, Chon Y, Huang XY, Lin SL. Patient-derived joint counts are a potential alternative for determining Disease Activity Score. The Journal of rheumatology. 2010;37(5):1035-41.
13. Levy G, Cheetham C, Cheatwood A, Burchette R. Validation of patient-reported joint counts in rheumatoid arthritis and the role of training. The Journal of rheumatology. 2007;34(6):1261-5.
14. Radner H, Grisar J, Smolen JS, Stamm T, Aletaha D. Value of self-performed joint counts in rheumatoid arthritis patients near remission. Arthritis research & therapy. 2012;14(2):R61.
15. Salaffi F, Filippucci E, Carotti M, Naredo E, Meenagh G, Ciapetti A, et al. Inter-observer agreement of standard joint counts in early rheumatoid arthritis: a comparison with grey scale ultrasonography--a preliminary study. Rheumatology (Oxford, England). 2008;47(1):54-8.
16. Uhlig T, Kvien TK, Pincus T. Test-retest reliability of disease activity core set measures and indices in rheumatoid arthritis. Annals of the rheumatic diseases. 2009;68(6):972-5.
17. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychological bulletin. 1979;86(2):420-8.

18. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychological assessment. 1994;6(4):284.
19. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. Journal of strength and conditioning research. 2005;19(1):231-40.
20. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet (London, England). 1986;1(8476):307-10.
21. Cohen J. A coefficient of agreement for nominal scales. Educational and psychological measurement. 1960;20(1):37-46.
22. Landis JR, Koch GG. The measurement of observer agreement for categorical data. biometrics. 1977:159-74.
23. Team RC. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2014. 2014.
24. Revelle W. psych: Procedures for Psychological, Psychometric, and Personality Research [Software]. 2016.